Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 46 (2015) 1668 – 1675

Procedia Computer Science

International Conference on Information and Communication Technologies (ICICT 2014)

A Simple and Robust Colour Based Video Copy Detection on Summarized Videos

Renu Mary Thomas [a,]*, Sumesh M S[a]

[a]*Adi Shankara Institute of Engineering and Technology, Kalady,Ernakulam, 683574, India .*

Abstract

In the digital world, all documents are accessible to the users in digital form and there are chances of it being copied or manipulated .This causes the violation of exclusive right of the content provider to publish the document. This call for the evolution of a copy detection system that determine whether a video has undergone any content preserving operation. This paper accentuates on the content-based technology using color correlation on summarized videos and also incorporating temporal features for matching the video sequences. This method is capable of handling videos with different frame rate without processing video on the whole.

Peer-review under responsibility of organizing committee of the International Conference on Information and Communication Technologies (ICICT 2014)

Keywords: Color correlation;Video summarization;TIRIs

1. Introduction

Copy detection refers to a method to identify whether a multimedia object is a copy of an already existing object. Multimedia objects can be audio, video or images. Considering the case of videos, many numbers of videos are uploaded to the internet every day or even every second. These may be distorted or manipulated videos

* Corresponding author. Tel.: +91-949-569-1190

E-mail address: renumerythomas@gmail.com

Peer-review under responsibility of organizing committee of the International Conference on Information and Communication Technologies (ICICT 2014)
doi:10.1016/j.procs.2015.02.106

of the already existing ones. This is an issue of great threat to the providers of the original content. This refers to the copyright infringement and it requires an efficient, fast and accurate video copy detection system. Websites such as YouTube which shares multimedia objects are current applications of this system.

The video copy detection can be performed using three main techniques. They are the hash based technique, watermarking technique and content based fingerprinting technique. Of these the hash based technique checks only for the absolute equality of the video and the technique may fail in the case of perceptually identical videos. A watermarking technique[1, 2] needs to add extra data into the existing video data and is not applicable to videos once distributed .So we use content based fingerprinting technique. Content based fingerprinting is the process of extracting fingerprints or compact signatures that uniquely identifies content of a video regardless of any content preserving distortions.

1.1. Properties of Fingerprint

Ideally, the content based fingerprint must be

- *Robust*: A video fingerprint should remain same even though the videos undergo some content preserving operations.
- *Discriminating*: The video fingerprints must differentiate between two different videos if they are distinctly different.
- *Compact*: A video fingerprint should be compact in size, comparing to the data size of the original video content.
- *Fewer complexes*: The method for extracting the fingerprint must be less complex.
- *Efficient for matching and search*: The fingerprints once extracted must be efficient for carrying out any sort of matching and searching operations.

1.2. Objective

The purpose of the work here is to develop an effective method for video copy detection with simple features, satisfactory discriminability, resistance to content-preserving operations ,satisfactory time and space complexity that works on a large variety of videos with different fps, frame size, resolution, different color map.

Color signatures are among the first being used in video fingerprinting. It is the simplest feature extraction technique than spatial, temporal or transform-domain signatures. A detailed study on the color based signatures reveal that absolute color[5-12] does not effectively deals with rotation, flipping and shifting operations and also does not effectively resist above said content preserving operations. Moreover these techniques[9] are found to be expensive and difficult.

In this paper a color correlation based video copy detection which works on a much summarized form of video is developed. Here the input video is divided into shots. The video is then processed in shot level and its TOC should be generated. Selecting one shot from each group of all scenes, a TIRI image is constructed. Its output is first transformed into R, G and B channels and is then divided into bxb blocks. Color correlation is then extracted and the percentage of number of pixels belonging to a particular group is calculated and is normalized to obtain the feature set .Color correlation histogram can be plotted using the feature vectors. This process is done for both the original video and the query video. Thus two feature sets are generated and the distance measure between these two identifies whether the query video is a copy of the already existing video.

From the theoretical analysis and experimental results, it can be clearly observed that this technique can effectively resist most common content-preserving operations, particularly rotation, flipping etc.

1.3. Application and Scope

Multimedia sharing web-sites such as YouTube, Facebook where thousands of videos are being shared in every second is the current application of the video copy detection system. Today, all of major Hollywood film and TV studios have adopted video fingerprinting technology. There are numerous researches going on in this field to avoid copyright infringement.

2. Overview

In the proposed system input video is divided into shots. The video is then processed in shot level and its TOC should be generated. Selecting one shot each from the summarized form a TIRI image is constructed. Its output is first transformed into R, G and B channels and is then divided into bxb blocks. Color correlation is then extracted and the percentage of number of pixels belonging to a particular group is calculated and is normalized to obtain the feature set .Color correlation histogram can then be plotted using the feature vectors. This process is done for both the original video and the query video. Thus two feature sets are generated and the distance measure between these two identifies whether the query video is a copy of the already existing video.

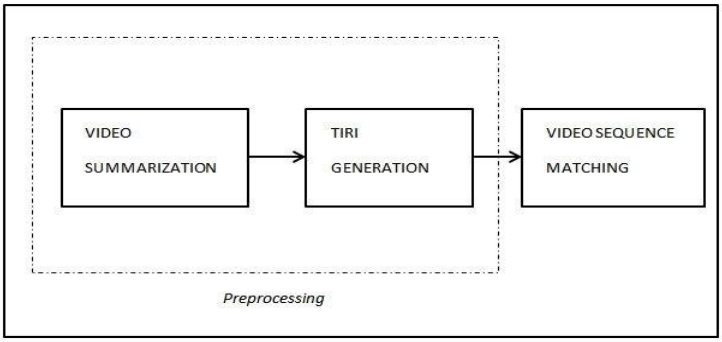

Fig. 1. Overall system

3. Methodology

3.1. Video Summarization

Video summarization[18] is based on TOC based summarization method. TOC stands for Table of Contents .It resembles the table of contents of a book. Using TOC of a book a reader can find out the chapter that suites his need without reading the whole book. Likewise TOC of video helps in easy browsing, accessing and retrieval of video contents. In the summarized form a video is divided into scenes which are again divided into groups and then to shots based on the similarity measure. A brief version of the algorithm[18] that constructs a scene based representation is as follows:

- Initialize.
- Assign first shot to first group and first scene.
- Get next shot.
- Test if current shot can be merged to an existing group. If yes, merge else form a new group.
- Test if shot i can be merged to an existing scene. If yes, merge else form a new scene.
- Goto step 2

Structural Similarity

The process of merging a current shot to an existing shot is based on the structural similarity measure. The structural similarity (SSIM) index measures the similarity between two images. The method works by comparing an image with a reference image .This reference image is an original, manipulation free image. SSIM outperforms PSNR and MSE. Structural similarity works with the idea that when the pixels are inter-related then they will be spatially close. This gives us idea about the structure of objects in the scene of a video. The SSIM[18] is calculated on different blocks of an image. The measure between two windows x and y is:

$$\mathrm{SSIM(X, Y)} = \frac{(2\mu_x\mu_y + C1)(2\sigma_{xy} + C2)}{(\mu^2{}_x + \mu^2{}_y + C1)(\sigma^2{}_x + \sigma^2{}_y + C2)} \quad (1)$$

where μ stands for the average, σ for variance, c1 and c2 stabilizing variables, σ_{xy} is the covariance. The value obtained by applying this formula will be a decimal value and it will be between -1 and 1. The structurally similar data will give a resultant 1.It is better to select the window of size 8x8.

3.2. Generation of TIRIs

TIRI[3] stands for temporal informative representative image .TIRI image is a single blurred image which contains all the motion information (temporal information) of the entire frames within the video .Such a representative image is formed by calculating the weighted average of all the frames. A TIRI[3] is obtained as

$$1'_{m,n} = \sum^{J}{}_{k=1}\, w_k l_{m,n,k} \quad (2)$$

where a TIRI image is generated by multiplying the weight w by the luminance of (m,n)th pixel of the k th frame of a set of J frames. The value of w_k is γ^k where the value of γ changes from 0 to 1 (exponential weighing function is used.)

3.3. Video Sequence matching

3.3.1. Color Correlation

Color correlation[17] is the organization of red, green and blue components of the RGB color space in accordance with their intensity in a particular image. There are six possible combinations of R,G and B. Thus color correlation groups intensity of frames of a video into following six cases:

case #1: Rxy ≥ Gxy ≥ Bxy, case #2 : Rxy ≥ Bxy ≥ Gxy
case #3: Gxy ≥ Rxy ≥ Bxy , case #4 : Gxy ≥ Bxy ≥ Rxy
case #5: Bxy ≥ Rxy ≥ Gxy, case #6 : Bxy ≥ Gxy ≥ Rxy
where 1 ≤ x ≤ w, 1 ≤ y ≤ h.

3.3.2. Color Correlation Histograms

Color Histogram[17] is generated by distinctly classifying the colors in an image and then taking a count of the occurrences of each color. Here, after color correlation six groups which contain the number of occurrences in each cases are generated and then the histogram can be plotted

3.3.3. Video Sequence Matching

If H_q, H_t are the TIRIs generated then the distance d between them is given by[17]

$$\mathrm{d}(H_q, H_t) = {}' 1/c \sum^{6}{}_{i=1} |H_q(i) - H_t(i)| \quad (3)$$

where C=2,since the normalization is in the range of [0,1].
The value of d will be close to 0 if a match occurs between two videos and d will be far-off from 0 when there is no match. Thus d=0 gives a copy pair and the contents will be different for the two videos otherwise.

4. Result and discussion

In this experimental set up, a video database consisting of 22 videos are selected[16]. The videos to the dataset are chosen from the open video project. 70% of the selected videos are of same resolution. On each video eight different distortions are performed to generate a query video. The distortions or the content preserving operations carried out here are letter box, pillar box, cropping, and insertion of pattern, rotation, flipping, Gaussian noise and picture in picture. The distance measure between the original video and the query video is then calculated to check for a match between the two videos. Table. 1. shows the distance measures.

Table 1. Distance measure

Videos	Letterbox	Pillar-box	Crop	Pattern	Rotate	Flip	G. Noise	Pic in Pic
V1	0.01	0.11	0.12	0.10	0.08	0.07	0.11	0.17
V2	0.20	0.18	0.20	0.20	0.12	0.18	0.14	0.15
V3	0.07	0.07	0.10	0.07	0.21	0.01	0.16	0.08
V4	0.02	0.02	0.03	0.01	0.18	0.00	0.01	0.03
V5	0.02	0.01	0.01	0.02	0.16	0.02	0.16	0.03
V6	0.09	0.09	0.14	0.06	0.20	0.21	0.16	0.11
V7	0.03	0.02	0.21	0.02	0.18	0.01	0.19	0.01
V8	0.31	0.16	0.33	0.03	0.31	0.11	0.33	0.02
V9	0.02	0.05	0.04	0.02	0.02	0.02	0.04	0.02
V10	0.13	0.12	0.12	0.13	0.13	0.11	0.12	0.11
V11	0.01	0.14	0.14	0.05	0.01	0.03	0.14	0.07
V12	0.00	0.02	0.02	0.02	0.00	0.01	0.02	0.04
V13	0.00	0.09	0.09	0.03	0.00	0.10	0.02	0.04
V14	0.06	0.03	0.03	0.07	0.06	0.16	0.03	0.08
V15	0.02	0.03	0.02	0.03	0.02	0.02	0.12	0.18
V16	0.13	0.02	0.02	0.01	0.13	0.13	0.02	0.07
V17	0.01	0.13	0.11	0.14	0.01	0.01	0.11	0.02
V18	0.01	0.05	0.07	0.21	0.01	0.01	0.12	0.01
V19	0.10	0.02	0.04	0.33	0.10	0.10	0.04	0.09
V20	0.06	0.03	0.04	0.04	0.06	0.06	0.04	0.02
V21	0.05	0.02	0.02	0.06	0.05	0.05	0.29	0.07
V22	0.04	0.03	0.03	0.05	0.04	0.04	0.12	0.06

The table gives the distance between an original video and the query video. From the distance values it can be seen that the similarity distances for copy pairs are close to 0. For video frames with different contents the distances would be far from 0. The picture in picture and insertion of pattern is primarily dependent on the color correlation of the video when a picture or pattern is inserted.

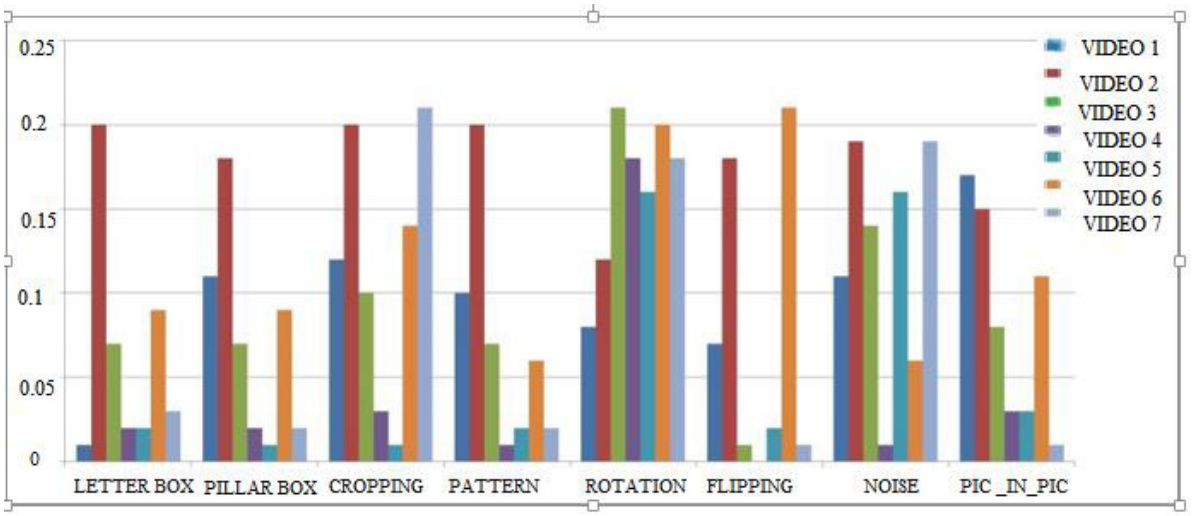

Fig. 2 . Distance measure graph.

The graph above plots the distance measure. From the graph it is apt to set the threshold value as .25. Any distance measure below this threshold gives a match whereas any value above the threshold shows that the two videos does not match. Table. 2. gives the TPR and FPR. TPR stands for the true positive rate , FPR stands for false positive rate and it is given by

TPR= No of correctly classified copies ÷ No of correct copies.

FPR= No of videos misclassified as copies ÷ No of noncopies

Table 2. TPR and FPR

Distortion	TPR	FPR
Letter box	0.95	0.016
Pillar box	1.00	0.01
Cropping	0.95	0.024
Insertion of pattern	0.95	0.016
Rotation	0.95	0.02
Flipping	1.00	0.16
Gaussian Noise	0.90	0.16
Picture in picture	1.00	0.01

Higher TPR means higher robustness and smaller FPR means good discriminability capabilility.So the ratios here show that the proposed system has good robustness and discriminabily characteristic . The Fig.3 below shows the maximum percentage of distortions that could be performed on a video and could still show a match with the original one.

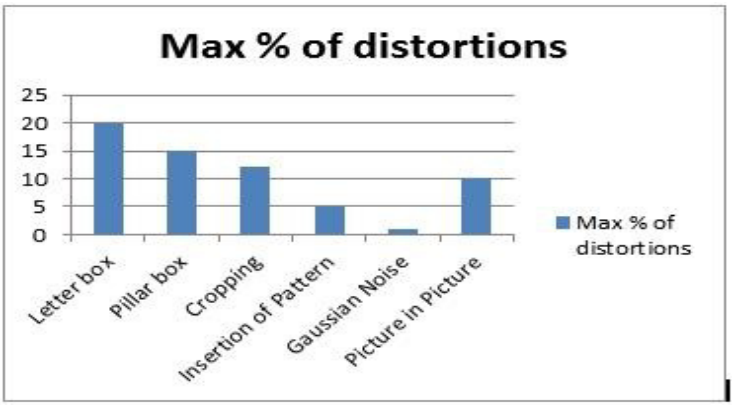

Fig. 3 . Maximum possible distortions

In the case of letterbox converting 20% from top and bottom boundary to 0 is the maximum withstandable distortion. For pillarbox 15% from left and right boundary, for cropping 12% from each boundary is the maximum distortion allowed.Inserting a pattern which is 5% of the original image and a picture which is 10% of the original image is also allowed.All types of rotation and flipping can also be effectively detected. The below given tables Table. 3. and Table. 4. gives a comparison of the proposed system with the system with no preprocessing stages such as summarization and TIRI generation.

Table 3. Comparison of TPR and FPR.

Distortion	Without preprocessing		With Preprocessing	
	TPR	FPR	TPR	FPR
Letter box	0.72	0.14	0.95	0.016
Pillar box	0.77	0.11	1.00	0.01
Cropping	0.72	0.06	0.95	0.024
Insertion of pattern	0.72	0.11	0.95	0.016
Rotation	0.77	0.10	0.95	0.02
Flipping	0.77	0.16	1.00	0.16
Gaussian Noise	0.68	0.11	0.90	0.16
Picture in picture	0.77	0.06	1.00	0.01

Table 4. Average Execution Time

Method	Avg. Execution Time
Video copy detection without preprocessing	34 s
Video copy detection with preprocessing	30 s

The Table. 4. gives the execution time while running the proposed system in MATLAB R2013a , Intel Core i5 processor(2.40 GHz),4 GB memory.

While comparing the proposed work with the already existing techniques (Table. 5.) it is clear that the proposed techniques has better quality because it correctly classified more number of videos.The better quality is due to the system's capability of handling videos with different resolution and frame rate .

Table 5. Comparison with the existing work

Existing Methods	Spatial	Temporal	SIFT	Color hist	Proposed System
No:of correctly classified videos	13/22	14/22	15/22	15/22	21/22

It is clear from Table. 3,Table. 4 and Table. 5 that even though video summarization and TIRI generation are both tedious tasks , the system shows better results when these tasks are performed because the system doesnot need to process the video on the whole and system becomes capable of handling videos with different frame rate, frame size and those with variations in color map.

5. Conclusion

A fast and robust video copy detection system, without processing video on the whole and using simple features can still achieve better discriminability and higher quality. The results show the weakness of the method against color correlation changes. To strengthen this method, generate an acoustic fingerprint and combine it with the visual fingerprint as future work. The matching is performed using the weighted average of the matching values of visual fingerprints and acoustic fingerprints.

References

1. F. Hartung and M. Kutter. Multimedia watermarking techniques. *Proc. IEEE* **87**: 7. p. 1079–1107; Jul 1999.
2. Y. Li and RH Deng. Publicly verifiable ownership protection for relational databases. *Proc. ACM Symp. Inform. Comput. Commun. Security*; 2006. p. 78–89.
3. Mani Malek Esmaeili, Mehrdad Fatourechi, and Rabab Kreidieh Ward. A Robust and Fast Video Copy Detection System Using Content Based Fingerprinting. *IEEE Trans on Information Forensics And Security* **6**: 1;Mar 2011
4. J Lu, EJ Delp, J Dittmann, ND Memon, and PW Wong.Video fingerprinting for copy identification: From research to industry applications. *Eds SPIE* 7254: 1.2009. p. 725402
5. Hampapur A, Hyun KH and Bolle RM. Comparison of sequence matching techniques for video copy detection. *Proc. SPIE, Storage and Retrieval for Media Databases.* 4676.p. 194-201.;Jan 2002.
6. C. Kim and B. Vasudev. Spatiotemporal sequence matching for efficient video copy detection. *IEEE Trans. Circuits Syst. Video Technol* 15:1. p. 127–132.;Jan 2005.
7. L. Chen and F. Stentiford. Video sequence matching based on temporal ordinal measurement. *Patter Recognit. Lett*.**29**:13. p. 1824– 1831; Oct. 2008.
8. XiaWu,Chong Wab,Alexander G.Hauptmann. Real time near duplicate elimination for video search with content and context. *IEEE Trans;*2009
9. L. Shang, L. Yang, F. Wang, KP Chan and XS Hua. Real-time large scale near-duplicate web video retrieval.*Proc. ACM Int. Conf. Multimedia*; 2010. p. 531–540.
10. Roopalakshmi R. Compact and efficient CBCD scheme based on integrated color features. *IEEE Int. Conf on Recent Trends in Information Technology;* 2011.
11. Gupta v,Varcheie,Gagnon,Boulianne . Content based video copy detection using nearest neighbour mapping. *IEEE Int. Conf. Information Science, Signal Processing and their Applications ;*2012.
12. Hong Liu, Hong Lu, and Xiangyang Xue. A Segmentation and Graph-Based Video Sequence Matching Method for Video Copy Detection .*IEEE Trans On Knowledge and Data Engineering*,**25**: 8;Aug 2013
13. P Brasnett, S Paschalakis, and M Bober. Recent developments on standardization of MPEG-7 visual signature tools. *Proc. IEEE Int. Conf. Multimedia Expo*; Sep. 2010.p. 1347–1352.
14. MJ Swain and D H Ballard. Color indexing. *Int. J. Comput. Vision*.**7**: 1. p. 11–32; Jun. 1991.
15. J Law-To, A Joly and N Boujemaa. 2007. Muscle-VCD-2007: *A Live Benchmark for Video Copy Detection.* Available Online.
16. The Open Video Project. *A Shared Digital Video Collection* .Available Online
17. Yanqiang Lei, Weiqi Luo, Yuangen Wang, Jiwu Huang, Video Sequence Matching Based on the Invariance of Color Correlation.*IEEE Trans on Circuits and Systems for Video Technology*..**22**: 9;Sep 2012
18. Yong Ruiy, Ziyou Xiongy, Regunathan Radhakrishnanz, Ajay Divakaranz, Thomas S. Huangy.A Unified Framework for Video Summarization, Browsing and Retrieval.*IEEE Int. Conf on Multimedia Signal Processing* ;2009